An L^1 Counting Problem in Ergodic Theory

Idris Assani, Department of Mathematics,
University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27599, USA
email: assani@email.unc.edu
www.math.unc.edu/Faculty/assani

Zoltán Buczolich,* Department of Analysis, Eötvös Loránd
University, Pázmány Péter Sétány 1/c, 1117 Budapest, Hungary
email: buczo@cs.elte.hu
www.cs.elte.hu/~buczo

and

R. Daniel Mauldin,† Department of Mathematics,
University of North Texas, Denton, Texas 76203-1430, USA
email: mauldin@unt.edu
www.math.unt.edu/~mauldin

November 1, 2018

*This work was completed while this author visited the Departments of Mathematics of University of North Texas and of University of North Carolina at Chapel Hill.

†Supported in part by NSF grant DMS 0100078

2000 Mathematics Subject Classification: Primary 37A05; Secondary 28D05, 47A35, 60F99.

Keywords: ergodic theorem, weak maximal inequality, return time theorems

Abstract

We solve the following counting problem for measure preserving transformations. For $f \in L^1_+(\mu)$, is it true that $\sup_n \dfrac{\mathbf{N}_n(f)(x)}{n} < \infty$, where
$$\mathbf{N}_n(f)(x) = \#\left\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\right\}?$$
One of the consequences is the nonvalidity of J. Bourgain's Return Time Theorem for pairs of (L^1, L^1) functions.

1 Introduction

Let $(X, \mathcal{B}, \mu)$ be a probability measure space, T an invertible measure preserving transformation on this space and $f \in L^1_+(\mu)$. Since $\frac{f(T^n x)}{n} \to 0$ a.e., the following function

$$\mathbf{N}_n(f)(x) = \#\left\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\right\}$$

is finite a.e. In this paper we consider the following

Counting Problem I. *Given* $f \in L^1_+(\mu)$ *do we have* $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} < \infty$, μ *a.e.?*

In [1] and [2] the operator $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ was introduced and the pointwise convergence of $\frac{\mathbf{N}_n(f)(x)}{n}$ was studied. It was shown there that if $f \in L^p_+$ for $p > 1$, or $f \in L \log L$ and the transformation T is ergodic, then $\frac{\mathbf{N}_n(f)(x)}{n}$ converges a.e to $\int f d\mu$. If T is not ergodic, then the limit is the conditional expectation of the function f with respect to the σ field of the invariant sets for T. Hence, the limit is the same as the limit of the ergodic averages $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)$. The limit of the ergodic averages, by Birkhoff's pointwise ergodic theorem, exists for any function $f \in L^1(\mu)$. It is natural to ask whether $\frac{\mathbf{N}_n(f)(x)}{n}$ also converges a.e., when $f \in L^1(\mu)$. Another motivation for this question is given by the fact that for i.i.d. random variables $X_n \in L^1$ it was shown in [1] that

$$\frac{\#\{k : \frac{X_k(\omega)}{k} > \frac{1}{n}\}}{n}$$

converges a.e. to $E(X_1)$. The counting problem was afterwards discussed in [9].

One can see by using the methods of [1], for instance, that the convergence for all functions $f \in L^1_+(\mu)$ will be guaranteed if one can answer the following equivalent problem.

Counting Problem II. *Does there exist a finite positive constant C such that for all measure preserving systems and all $\lambda > 0$*

$$\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} \leq \frac{C}{\lambda}\|f\|_1?$$

Our main result will be to show that this equivalent problem has a negative answer. More precisely we have

Theorem 1.

$$\sup_{(X,\mathcal{B},\mu,T)} \sup_{\|f\|_1=1} \sup_{\lambda>0} \lambda \cdot \mu \left\{ x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda \right\} = \infty.$$

This theorem answers then the question raised in [1].

We will also derive answers to some related problems. The first consequence, linked to the study of the maximal function $\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, is what we call the return times for the tail (of the Cesaro averages).

Definition 1. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The Return Times for the Tail Property holds in $L^r(\mu)$, $1 \leq r \leq \infty$ if for each $f \in L^r(\mu)$ we can find a set X_f of full measure such that for all $x \in X_f$ for all measure preserving systems $(Y, \mathcal{G}, \nu, S)$ and each $g \in L^1(\nu)$ the sequence $\frac{f(T^n x) \cdot g(S^n y)}{n}$ converges to zero for a.e. y.

A first consequence of Theorem 1 will be the following

Theorem 2. *The Return Times for the Tail Property does not hold for* $p = 1$.

We observe that in [1] and [2] it was shown that the Return Times for the Tail Property holds in L^p for $1 < p \leq \infty$ and even in $L \log L$.

A second consequence is a solution to the (L^1, L^1) problem mentioned in [1], [3] and [14]. To explain this problem we need a few definitions.

Definition 2. A sequence of scalars a_n is said to be good universal for the pointwise ergodic theorem (resp. norm convergence) in L^r, $1 \leq r \leq \infty$ if for all dynamical systems $(Y, \mathcal{G}, \nu, S)$ the averages

$$\frac{1}{n} \sum_{k=1}^{n} a_k \cdot g(S^k y)$$

converge a.e. (resp. in $L^r(\nu)$ norm).

In [4], [5], and [6] J. Bourgain showed that given $f \in L^\infty(\mu)$ the sequence $f(T^n x)$ is μ a.e. good universal for the pointwise convergence in L^1. Using Hölder's inequality and the maximal inequality for the ergodic averages one can extend his result to the pairs (L^p, L^q) where $\frac{1}{p} + \frac{1}{q} = 1$. This was mentioned in [13]. Bourgain's Return Time Theorem strengthens Birkhoff's theorem on the product space when the functions, f and g, respect duality. That is, if the function $f \in L^p(\mu)$ for some $1 \leq p \leq \infty$, then the set of convergence obtained from the Return Times Theorem works for all functions $g \in L^q(\nu)$, where $\frac{1}{p} + \frac{1}{q} = 1$, hence it is a universal set. However, fixing f and g, the projection of the convergence set onto the first factor obtained by Birkhoff's theorem depends on both functions. A weakness of the Return Time Theorem is that it does not address the case of $f \in L^1$ and $g \in L^1$. Birkhoff's theorem, on the other hand, guarantees convergence for $f \otimes g \in L^1 \times L^1$, $\mu \otimes \nu$-almost everywhere.

In [3] random stationary weights (i.i.d. random variables) were given for which one could go "beyond" the duality apparently imposed by the use of

Hölder's inequality. It was also shown that given $f \in L^1(\mu)$ the sequence $(f(T^n x))$ is μ-a.e. good universal for the L^1 norm. In [1] a Multiple Return Times Theorem for L^1 i.i.d. random variables was obtained while in [14] a Multiple Return Times theorem was proved for L^∞ stationary processes. The (L^1, L^1) problem was the following.

(L^1, L^1) **Problem.** *Given $f \in L^1(\mu)$, is the sequence $(f(T^n x))$, μ-a.e. good universal for the pointwise ergodic theorem in L^1?*

A consequence of Theorem 2 is the following solution to the (L^1, L^1) problem

Theorem 3. *Bourgain's Return Time Theorem does not hold for pairs of (L^1, L^1) functions.*

We also derive in Section 4 some consequences in $L^1(\mathbb{T})$ between the continuous analog of the maximal function $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, namely

$$A(f)(x) = \sup_t t \cdot m \left\{ 0 < y < x : \frac{f(x-y)}{y} > t \right\},$$

or, analogously,

$$A(f)(x) = \sup_t t \cdot m \left\{ 0 < y < x : \frac{f(y)}{x-y} > t \right\},$$

and the one sided Hardy–Littlewood maximal function.

2 Proof of Theorem 1

In this section μ will denote Lebesgue measure on $\mathbb{R}$ and log will denote logarithm in *base* 2. An interval I is a 2^{-R} grid interval if there is some $j \in \mathbb{Z}$ such that $I = [j \cdot 2^{-R}, (j+1)2^{-R})$.

2.1 Basic systems

A "life" function is a map $\nu : \mathbb{N} \to \mathbb{N}$ such that for each $N \in \mathbb{N}$, $\nu(N) > N$. Given a life function ν, a gain constant $M > 3$, and a startup time N_1 we choose a sequence $N_2, ..., N_M$ so that

$$N_l = 20 + \nu(N_{l-1}), \ l = 2, ..., M. \tag{1}$$

.

Our aim in this section is to prove the following

Lemma 4. *Suppose that a gain constant $M > 3$, a life function ν, a support constant $S < 2^M$, and a startup time $N_1 > \max\{10, M\}$ are given. Choose the sequence $N_2, ..., N_M$ based on M, ν, and N_1 satisfying (1). Given any 2^{-R} grid interval I, there exists a positive integer $J_0 > R$, disjoint subsets $\Gamma_1, ..., \Gamma_M$ of I, and for each integer $J \geq J_0 > R$ there is a simple function $f : \mathbb{R} \to \mathbb{R}$, such that $f(x) = 0$ for $x \notin I$ and if $T(x) = x + 2^{-J}$ then for all*

$l = 1, ..., M$,

$$\frac{\mathbf{N}_n(f)(x)}{n} > 0.99 \cdot 2^{-l+1} \qquad \textit{when } 2^{N_l} \leq n \leq 2^{\nu(N_l)} \tag{2}$$

for all $x \in \Gamma_l$. *Moreover, each set* Γ_l *consists of the union of intervals of the form* $[i \cdot 2^{-J_0}, (i+1)2^{-J_0})$, $\mu(\Gamma_l) > 0.99 \cdot 2^{-M+l-1}\mu(I)$, *and* $\int_I f = 2^{-M+1}\mu(I)$. *We can also require that* $f(x) = 0$ *for any* x *which is not in an interval of the form* $[(i \cdot 2^M + S)2^{-J}, (i \cdot 2^M + S + 1)2^{-J})$ *for some* $i \in \mathbb{Z}$.

Proof. Set $h_0 = 2^{M+10}$ and choose J_0 such that

$$2^{10}2^{\nu(N_M)}h_0 2^{-J_0} < 2^{-R}, \tag{3}$$

or equivalently, $J_0 > \nu(N_M) + M + 20 + R$. Now, let an integer $J \geq J_0$ be given. Set $h = h_0 \cdot 2^{J-J_0} = 2^{M+10+J-J_0}$. We shall first define a sequence of sets $B_M, B_{M-1}, \ldots, B_1$ each as the union of some intervals in a corresponding sequence of finer dyadic grids. To begin put

$$\begin{aligned} B_M =& I \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J} \cdot 2^{N_M} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_M} h) \\ =& I \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J_0} \cdot 2^{N_M} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_M} h_0). \end{aligned} \tag{4}$$

Thus, B_M consists of the intervals in the standard $2^{N_M+M-J_0}$ grid with even index, j and that are subsets of the interval I. Clearly, $\mu(B_M) = \mu(I)/2$.

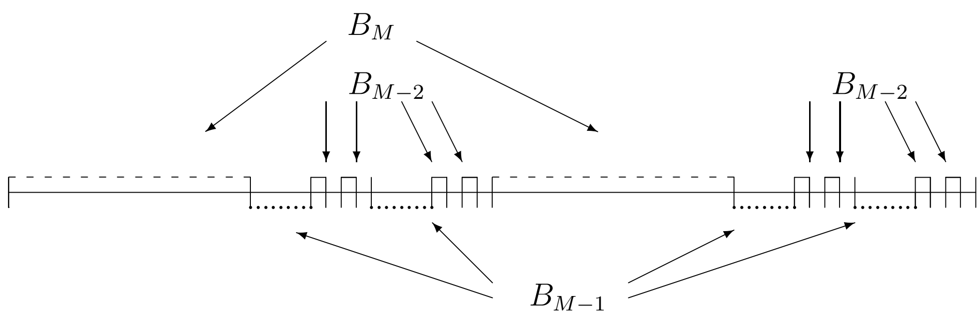

Figure 1: The sets B_M, B_{M-1}, and B_{M-2}

In Figure 1 we illustrate the manner in which the sets B_l, $l = M-2, M-1, M$, are located in I. Of course, in an illustration we cannot divide an interval into several thousand pieces, so in the figure the set B_M consists of two intervals of length $\mu(I)/4$, marked by dashed line, B_{M-1} consists of four intervals of length $\mu(I)/16$, marked by dotted line, B_{M-2} consists of eight intervals marked by solid lines. The complement of $B_M \cup B_{M-1} \cup B_{M-2}$ consists of eight "unmarked" intervals, each of the same length as the components of B_{M-2}.

In (4) the first expression for B_M is given for some computational purposes whereas the second expression shows that B_M does not depend on J but rather on J_0. The same is true for all the sets B_i to be defined now.

Assume that $l \in \{0, ..., M-3\}$ and $B_{M-l'}$ is given for all $l' \in \{0, ..., l\}$.

Set

$$B_{M-(l+1)} = \tag{5}$$
$$= (I \setminus \bigcup_{l'=0}^{l} B_{M-l'}) \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J} \cdot 2^{N_{M-l-1}} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_{M-l-1}} h) =$$
$$= (I \setminus \bigcup_{l'=0}^{l} B_{M-l'}) \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J_0} \cdot 2^{N_{M-l-1}} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_{M-l-1}} h_0).$$

Thus, the set $B_{M-(l+1)}$ consists of the intervals with even index in the standard $2^{N_{M-l-1}+M-J_0}$ grid that are subsets of I and are not in $\cup_{i=M-l}^{M} B_i$.

Finally, if B_{M-l} is given for $l \in \{0, ..., M-2\}$, we set

$$B_1 = B_{M-((M-2)+1)} = I \setminus \bigcup_{l=0}^{M-2} B_{M-l}.$$

Returning to the illustration on Figure 1, if $M = 4$ then $B_M = B_4$ is marked by the dashed line, B_3 is by the dotted line, B_2 by the solid line, and B_1, the complement of the other three is the "unmarked" part of I.

Observe that $\mu(B_{M-l}) = \mu(I)/2^{l+1}$ holds for $l = 0, ..., M-2$ and $\mu(B_1) = \mu(I)/2^{M-1} > \mu(I)/2^M = \mu(I)/2^{(M-1)+1}$. The set B_1 is the union of some disjoint intervals of the form

$$[(2j-1) \cdot 2^{-10-J} \cdot 2^{N_2} h, 2j \cdot 2^{-10-J} \cdot 2^{N_2} h) = \tag{6}$$
$$= [(2j-1) \cdot 2^{-10-J_0} \cdot 2^{N_2} h_0, 2j \cdot 2^{-10-J_0} \cdot 2^{N_2} h_0),$$

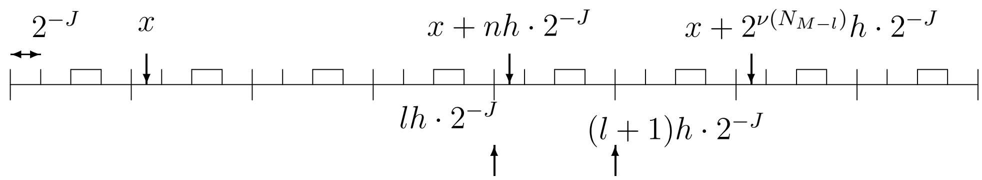

Figure 2: The definition of f in an interval I'

while for any $l = 0, ..., M-2$ the set B_{M-l} is the union of some intervals of the form

$$[2j \cdot 2^{-10-J} \cdot 2^{N_{M-l}} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_{M-l}} h) \tag{7}$$
$$[2j \cdot 2^{-10-J_0} \cdot 2^{N_{M-l}} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_{M-l}} h_0).$$

Our function f which depends on J will have value 0 on $\cup_{l=2}^{M} B_l$. To determine its values on B_1, consider one of the intervals making up B_1:

$$I' = [(2j-1) \cdot 2^{-10-J} \cdot 2^{N_2} h, 2j \cdot 2^{-10-J} \cdot 2^{N_2} h) \subset B_1.$$

For each $l \in \mathbb{Z}$ such that the interval $[lh \cdot 2^{-J}, (l+1)h \cdot 2^{-J}) \subset I'$ (and there are $\frac{2^{N_2}}{2^{10}}$ such l), choose exactly one l' such that $lh \leq l' < (l+1)h$, $l' \equiv S$ modulo 2^M and set $f(x) = h$ for $x \in [l' \cdot 2^{-J}, (l'+1) \cdot 2^{-J})$, otherwise we set $f(x) = 0$.

In Figure 2 one can see one interval I' being enlarged. Again we could not divide this interval in a drawing into several thousand subintervals, so in this illustration $h = 4$, and $S = 2$. One tiny interval is of length 2^{-J}, the tiny intervals marked by an extra solid line are the ones where $f = h$.

From the definition of f, we have $\int_{I'} f = \frac{h}{2^J} \cdot \frac{2^{N_2}}{2^{10}} = \mu(I')$. By summing this over all subintervals of B_1 of type I', we obtain $\int_I f = \int_{B_1} f = \mu(B_1) = \mu(I)/2^{M-1}$.

Suppose $2^{N_1} \leq n \leq 2^{\nu(N_1)}$, and

$$[x, x + h \cdot 2^{\nu(N_1)-J}) \subset I'. \tag{8}$$

Then $N_1 > 10$ implies $1000 \leq n$ and hence

$$\begin{aligned}\mathbf{N}_n(f)(x) =&\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\} = \\ =&\#\{k : hn > k \text{ and } f(T^k x) = h\} > 0.99 \cdot \frac{nh}{h} = 0.99n,\end{aligned}$$

Of course, instead of 0.99 we could have used 0.999, but this is not of any consequence for our purposes.

Now, we define the sets Γ_i which do not depend on J from the sets B_i. To begin set

$$\begin{aligned}\Gamma_1 =&\{x \in B_1 : [x, x + h \cdot 2^{\nu(N_1)-J}) \subset B_1\} = \\ =&\{x \in B_1 : [x, x + h_0 \cdot 2^{\nu(N_1)-J_0}) \subset B_1\}.\end{aligned} \tag{9}$$

Again, the second expression here shows that Γ_1 does not depend on J since

B_1 does not depend on J. For each interval I' making up B_1, by using (1), we have $|\Gamma_1 \cap I'| \geq |I'| - h \cdot 2^{\nu(N_1)-J} \geq |I'| \cdot (1 - 2^{-(N_2-\nu(N_1)-10)}) > 0.99|I'|$. So, $\mu(\Gamma_1) > 0.99 \cdot \mu(B_1)$.

Observe that for each $l = 1, ..., M-2$, the set $I \setminus \cup_{i=0}^{l-1} B_{M-i} = \cup_{i=1}^{M-l} B_i$ is the union of some intervals of the form

$$I'_{M-l} = [(2j-1) \cdot 2^{-10-J} \cdot 2^{N_{M-l+1}} h, 2j \cdot 2^{-10-J} \cdot 2^{N_{M-l+1}} h).$$

Also, the two sets B_{M-l} and $B_1 \cup \ldots \cup B_{M-l-1}$ are equally distributed in I'_{M-l} in the sense that if one takes the $2^{N_{M-l}} h / 2^{10+J}$ grid of the interval I'_{M-l}, then every evenly indexed interval is in B_{M-l} and the others are in $B_1 \cup \ldots \cup B_{M-l-1}$. In particular, $\mu(B_{M-l} \cap I'_{M-l}) = \mu(I'_{M-l})/2 = \mu(\cup_{i<M-l} B_i \cap I'_{M-l})$.

Finally, by induction one can also see that

$$\mu(B_1 \cap I'_{M-l}) = \mu(I'_{M-l})/2^{M-l-1}, \tag{10}$$

and, more generally, if $n \in [2^{N_{M-l}}, 2^{\nu(N_{M-l})}]$ and $[x, x + nh \cdot 2^{-J}) \subset I'_{M-l}$, then

$$\mu(B_1 \cap [x, x + nh \cdot 2^{-J})) > 0.995 nh \cdot 2^{-J} / 2^{M-l-1}. \tag{11}$$

Set

$$\Gamma_{M-l} = \{x \in B_{M-l} : [x, x + h \cdot 2^{\nu(N_{M-l})-J}) \subset \bigcup_{l' \leq M-l} B_{l'}\} = \tag{12}$$

$$= \{x \in B_{M-l} : [x, x + h_0 \cdot 2^{\nu(N_{M-l})-J_0}) \subset \bigcup_{l' \leq M-l} B_{l'}\}.$$

Using (1) one can see that $\mu(\Gamma_{M-l}) > 0.99\mu(B_{M-l}) \geq 0.99\mu(I)/2^{l+1}$. If $x \in \Gamma_{M-l}$ and I'_{M-l} is the subinterval of $\cup_{l'=1}^{M-l} B_{l'}$ containing x, then $x + jh \cdot 2^{-J} \in I'_{M-l}$ for all $0 \leq j \leq 2^{\nu(N_l)}$. By using (11) and the definition of $f(x)$ we have

$$\mathbf{N}_n(f)(x) = \#\{k : hn > k \text{ and } f(T^k x) = h\} \geq 0.99 \frac{nh}{h \cdot 2^{M-l-1}} = 0.99 \frac{n}{2^{M-l-1}}.$$

From $N_2 > N_1 > 10$, (4), (5), (6), (7), (9), and (12) it follows that each Γ_l is the union of intervals of the form $[i \cdot 2^{-J_0}, (i+1) \cdot 2^{-J_0})$.

□

2.2 Level k systems

In this section the gain constant $M \in \mathbb{N}$ is fixed.

Next we define the life functions for all $k \in \mathbb{N}$. We will use these functions in the proof of Lemma 5. Set $\nu_1(N) = N + 1$ for any $N \in \mathbb{N}$. We proceed by induction, so assume that for $k \in \mathbb{N}$ we have already defined ν_k. If some $N \in \mathbb{N}$ is given use $\nu = \nu_k$ and $N_1 = N_1^{(k)}(N) = N$ in (1) to determine the sequence $N_2^{(k)}(N), ..., N_M^{(k)}(N)$. Put $\nu_{k+1}(N) = \nu_k(N_M^{(k)}(N)) > N$.

We say that a random variable $X : I \to \mathbb{R}$ is $(M-0.99)$-distributed on I if $X(x) \in \{0, 0.99, 0.99 \cdot 2^{-1}, ..., 0.99 \cdot 2^{-M+1}\}$ and $\mu(\{x : X(x) = 0.99 \cdot 2^{-l+1}\}) = 0.99 \cdot 2^{-M+l-1}\mu(I)$, for $l = 1, ..., M$.

This section is about the existence of level k systems, by which we mean any system (T, f) satisfying the conditions described in the next lemma.

Lemma 5. *For any 2^{-R} grid interval I_0, positive integer $k \leq 2^M$, and any startup time $K_S^{(k)} > \max\{10, M\}$ there exists $J_0 > 0$ such that for all $J \geq J_0$ we can find a system (T, f) with the following properties. The transformation T is given by $T(x) = x + 2^{-J}$. We have independent $(M - 0.99)$-distributed random variables X_h, $h = 1, ..., k$, on I_0 and an exit time $K_e^{(k)}$ such that for any $x \in I_0$ there exists an $n \in [2^{K_S^{(k)}}, 2^{K_e^{(k)}}]$, for which*

$$\frac{\mathbf{N}_n(f)(x)}{n} \geq \sum_{h=1}^{k} X_h(x). \tag{13}$$

Moreover, f is constant on the intervals of the form $[i \cdot 2^{-J}, (i+1)2^{-J})$, $\int_{I_0} f = k \cdot 2^{-M+1}\mu(I_0)$, $f(x) = 0 = X_h(x)$ for $x \notin I_0$, $h = 1, ..., k$. We also may require that if

$$x \notin \bigcup_{l=0}^{k-1} \bigcup_{i \in \mathbb{Z}} [(i \cdot 2^M + l)2^{-J}, (i \cdot 2^M + l + 1) \cdot 2^{-J}),$$

then $f(x) = 0$.

Proof. To define our level 1 systems we use Lemma 4 on I_0. We apply Lemma 4 with $\nu = \nu_1$, and $N_1 = K_S^{(1)}$. So, $K_e^{(1)} = \nu_1(N_M)$ will be the exit time. We choose our $(M-0.99)$-distributed random variable the following way. For $l=1,...,M$ we select a measurable set $\widehat{\Gamma}_l \subset \Gamma_l$ such that $\mu(\widehat{\Gamma}_l) = 0.99 \cdot 2^{-M+l-1} \cdot 2^{-R}$. If $x \in \widehat{\Gamma}_l$ for some l then we set $X_1(x) = 0.99 \cdot 2^{-l+1}$ and $X_1(x) = 0$ otherwise. Viewed in this way Lemma 4 guarantees that level 1 systems exist.

We proceed by induction on k. Assume that level k systems exist and we need to verify the existence of level $k+1$ systems.

First, calling upon Lemma 4, we define a "mother" base system. The "subsystems" of this "mother" system will be level k systems with different life intervals. Here is a heuristic argument behind our construction. Due to the L^1 restrictions, the mother system is unable to deal with all the subsystems simultaneously at the same time. So some subsystems have longer and longer waiting times, but the longer the waiting time, the longer lifetime they need. Since we already know how the subsystems will look, this information is encoded by the life function ν_{k+1}. Now, using this function, we can "design" a mother system which can accomodate all the subsystems. Let us proceed.

Given the startup constant $N_1 = N_{1,0} = K_S^{(k+1)} > \max\{10, M\}$ putting the life function ν_{k+1} defined at the beginning of Subsection 2.2 into (1), determine the sequence $N_{2,0}, ..., N_{M,0}$, (the extra 0 in subscripts will refer to the "mother system"). We also put $N_{0,0} = N_1$, and set the support constant $S_0 = k$ for the mother system.

Next we apply Lemma 4 with $\nu = \nu_{k+1}$ to the 2^{-R} grid interval $I_0 =$

$[j_0 \cdot 2^{-R}, (j_0+1) \cdot 2^{-R})$ we choose $J_{0,0}$ and disjoint subsets $\Gamma_{1,0}, ..., \Gamma_{M,0}$ of I_0 such that for each $l = 1, ..., M$, $\Gamma_{l,0}$ consists of the union of some intervals of the form $[i \cdot 2^{-J_{0,0}}, (i+1)2^{-J_{0,0}})$, and $\mu(\Gamma_{l,0}) > 0.99 \cdot 2^{-M+l-1} \cdot 2^{-R}$. For any $J \geq J_{0,0}$ we can choose a function $\phi_0 = f : I_0 \to \mathbb{R}$, such that if $T(x) = x + 2^{-J}$ then for all $l = 1, ..., M$,

$$\frac{\mathbf{N}_n(\phi_0)(x)}{n} > 0.99 \cdot 2^{-l+1}, \quad \text{when } 2^{N_{l,0}} \leq n \leq 2^{\nu_{k+1}(N_{l,0})}, \tag{14}$$

for all $x \in \Gamma_{l,0}$. Moreover, $\int_{I_0} \phi_0 = 2^{-M+1}\mu(I_0)$. Since $S_0 = k$, we also have $\phi_0(x) = 0$ for any x which is not in an interval of the form $[(i \cdot 2^M + k)2^{-J}, ((i \cdot 2^M + k + 1)2^{-J})$ for some $i \in \mathbb{Z}$.

Next, consider the intervals $I_j = [j_0 \cdot 2^{-R} + (j-1) \cdot 2^{-J_{0,0}}, j_0 \cdot 2^{-R} + j \cdot 2^{-J_{0,0}})$ for $j = 1, ..., 2^{J_{0,0}-R}$. Our "subsystems" will live on these intervals.

If $I_j \subset \cup_{l=1}^{M} \Gamma_{l,0}$ then there is a unique $l(j)$ such that $I_j \subset \Gamma_{l(j),0}$. If $I_j \not\subset \cup_{l=1}^{M} \Gamma_{l,0}$ then $I_j \cap \cup_{l=1}^{M} \Gamma_{l,0} = \emptyset$, and in this case we set $l(j) = 0$. By our assumption on any I_j we can find level k systems. So, for each $j \in \{1, ..., 2^{J_{0,0}-R}\}$ choose a level k system on I_j with startup time $K_{S,j}^{(k)} = N_{l(j),0}$. Choose $J_{0,j}$ for each $j = 1, ..., 2^{J_{0,0}-R}$ according to our induction hypothesis. Set $J_0 = \max\{J_{0,j} : j = 0, ..., 2^{J_{0,0}-R}\}$ and choose a $J \geq J_0$. The transformation T will be given by $T(x) = x + 2^{-J}$. For this J choose ϕ_0 as was explained above, and by the induction hypothesis for any $j = 1, ..., 2^{J_{0,0}-R}$ choose $\phi_j = f$ and independent $(M - 0.99)$-distributed random variables $X_{h,j}$, $h = 1, ..., k$, on I_j, and an exit time $K_{e,j}^{(k)} = \nu_{k+1}(N_{l(j),0})$ such that for any $x \in I_j$ there exists an $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}]$, for which

$$\frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k} X_{h,j}(x). \tag{15}$$

Moreover, ϕ_j is constant on the intervals of the form $[i \cdot 2^{-J}, (i+1)2^{-J})$, $\int_{I_j} \phi_j = k \cdot 2^{-M+1}\mu(I_j)$, $\phi_j(x) = 0 = X_{h,j}(x)$, for $x \notin I_j$, $h = 1, ..., k$. We may also require that if

$$x \notin \bigcup_{l=0}^{k-1} \bigcup_{i \in \mathbb{Z}} [(i \cdot 2^M + l)2^{-J}, (i \cdot 2^M + l + 1) \cdot 2^{-J})$$

then $\phi_j(x) = 0$. This last property implies that the support of ϕ_0 is disjoint from the support of any ϕ_j, $j = 1, ..., 2^{J_{0,0}-R}$. Since ϕ_j is supported on I_j, we see that the supports of the functions ϕ_j are also disjoint.

Set $f = \sum_{j=0}^{2^{J_{0,0}-R}} \phi_j$. Then, using the fact that the supports are disjoint, we have $\mathbf{N}_n(f)(x) = \sum_{j=0}^{2^{J_{0,0}-R}} \mathbf{N}_n(\phi_j)(x)$. We also calculate

$$\begin{aligned}
\int_{I_0} f = \int_{I_0} \phi_0 + \sum_{j=1}^{2^{J_{0,0}-R}} \int_{I_j} \phi_j = \\
= 2^{-M+1}\mu(I_0) + k \cdot 2^{-M+1} \sum_{j=1}^{2^{J_{0,0}-R}} \mu(I_j) = (k+1)2^{-M+1}\mu(I_0).
\end{aligned}$$

For $h = 1, ..., k$, set $X_h = X'_h = \sum_{j=1}^{2^{J_{0,0}-R}} X_{h,j}$. Let $X'_{k+1}(x) = 0.99 \cdot 2^{-l+1}$ if $x \in \Gamma_{l,0}$, otherwise set $X'_{k+1} = 0$. Since X'_{k+1} is constant on the intervals I_j, one can also see that the functions $X'_h(x)$, $h = 1, ..., k+1$ are independent. The functions $X'_h(x)$ are $(M-0.99)$-distributed on I_0 for $h = 1, ..., k$. The function $X'_{k+1}(x)$ is not $(M-0.99)$-distributed, but is $(M-0.99)$-superdistributed. By this we mean that $\mu(\{x : X'_{k+1}(x) = 0.99 \cdot 2^{-l+1}\}) \geq 0.99 \cdot 2^{-M+l-1}\mu(I_0)$, for any $l = 1, ..., M$. But we can and do choose $X_{k+1} \leq X'_{k+1}$ such that X_{k+1} is $(M-0.99)$-distributed on I_0 and the system $X_h(x)$, $h = 1, ..., k+1$ is independent.

If $x \in I_j \subset \Gamma_{l(j),0}$, then

$$\frac{\mathbf{N}_n(\phi_0)(x)}{n} > 0.99 \cdot 2^{-l(j)+1} = X'_{k+1}(x) \geq X_{k+1}(x),$$

when $2^{N_{l(j),0}} \leq n \leq 2^{\nu_{k+1}(N_{l(j),0})}$. For these same x, by our induction hypothesis, there exists $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}]$ for which

$$\frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k} X_{h,j}(x) = \sum_{h=1}^{k} X_h(x).$$

Therefore, there exists $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}] \subset [2^{K_S^{(k+1)}}, 2^{\nu_{k+1}(N_{M,0})}]$ for which

$$\frac{\mathbf{N}_n(f)(x)}{n} = \sum_{j=0}^{2^{J_{0,0}-R}} \frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k+1} X_h(x).$$

This also shows that the exit time $K_e^{(k+1)}$ can be chosen to be $\nu_{k+1}(N_{M,0})$.

$\square$

2.3 p-blocks

We restate in our measure theoretical language formula (9) on p. 21 of [11] in the form of a lemma.

Lemma 6. *Assume that for a given $q \in \mathbb{N}$ we have independent identically distributed random variables $X_1, ..., X_q$ on a probability space (Ω, Σ, μ), each with finite mean u and variance v. Then for each $\epsilon > 0$ we have*

$$\mu\left(\left\{x : \left|\left(\sum_{h=1}^{q} X_h(x)\right) - qu\right| \geq q\epsilon\right\}\right) \leq \frac{qv}{(q\epsilon)^2}. \tag{16}$$

This section concerns the existence of p-blocks as described in the next lemma. We assume $I = [0, 1)$.

Lemma 7. *There exists $p_0 > 2$ such that for every $p > p_0$ we can choose a p-block. By this we mean, that we can find a system (T_p, f_p), such that $\frac{1}{p \log^2(p)} \leq \int_I f_p \leq \frac{4}{p \cdot \log^2(p)}$, $T_p(x) = x + 2^{-J_p}$, mod 1 for a large integer J_p. There is a set Λ_p with $\mu(\Lambda_p) > 0.99$ and there exists an exit time $E_p > 2^p$ such that for each $x \in \Lambda_p$ there is some $n \in [2^{2^p}, 2^{E_p}]$ for which*

$$\frac{\mathbf{N}_n(f_p)(x)}{n} > \frac{1}{2^2 \log^2(p)}.$$

Proof. Using Lemma 5 on $I_0 = I = [0, 1)$ with $k = 2^p$, $M = M_p = [p + \log(p) + \log(\log^2(p))]$, and startup time $K_S^{(2^p)} = 2^p$ we choose and fix $J_p \geq J_0$ and a level 2^p system (T_p, f_p) such that $T_p(x) = x + 2^{-J_p}$, mod 1. Here we remark that Lemma 5 uses $T_p(x) = x + 2^{-J_p}$, but f_p is supported on I and hence by using $T_p(x) = x + 2^{-J_p}$, mod 1 we cannot decrease $\mathbf{N}_n(f_p)$. We have independent $(M_p - 0.99)$-distributed random variables $X_{h,p}$, $h = 1, ..., 2^p$, such that for any $x \in I_0$ there exists $n \in [2^{2^p}, 2^{K_e^{(2^p)}}]$ for which

$$\frac{\mathbf{N}_n(f_p)(x)}{n} \geq \sum_{h=1}^{2^p} X_{h,p}(x),$$

and $\int_I f_p = 2^p \cdot 2^{-M_p+1}$.

Then, for any h,

$$\begin{aligned} u = \int_I X_{h,p} = & \sum_{l=1}^{M_p} 0.99 \cdot 2^{-l+1} \cdot 0.99 \cdot 2^{-M_p+l-1} \\ \geq & \frac{0.99^2}{2^p p \cdot \log^2(p)} \cdot p > \frac{1}{2^{p+1} \log^2(p)}. \end{aligned} \tag{17}$$

and

$$v_0 = \int_I X_{h,p}^2 = \sum_{l=1}^{M_p} 0.99^2 \cdot 2^{-2l+2} \cdot 0.99 \cdot 2^{-M_p+l-1} =$$
$$\sum_{l=1}^{M_p} 0.99^3 \cdot 2^{-M_p-l+1} < 0.99^3 \cdot 2^{-M_p+1} \sum_{l=1}^{\infty} 2^{-l} \leq \frac{4}{2^p \cdot p \cdot \log^2(p)}.$$

We have

$$0 < v = \int_I (X_{h,p}(x) - u)^2 dx = v_0 - u^2 < v_0 \leq \frac{4}{2^p \cdot p \cdot \log^2(p)}.$$

Next, by Lemma 6 used with $\epsilon = 1/2^{p+2}\log^2(p)$, $q = 2^p$ we obtain that

$$\mu\left(\left\{x : \left|\left(\sum_{h=1}^{2^p} X_{h,p}(x)\right) - 2^p u\right| \geq 2^p \cdot \frac{1}{2^{p+2}\log^2(p)}\right\}\right) \leq$$
$$\frac{2^p \cdot \frac{4}{2^p \cdot p \cdot \log^2(p)}}{(2^p \frac{1}{2^{p+2}\cdot\log^2(p)})^2} = \frac{64 \cdot \log^2(p)}{p}.$$

By using (17) this implies

$$\mu\left(\left\{x : \sum_{h=1}^{2^p} X_{h,p}(x) \leq \frac{1}{2^2\log^2(p)}\right\}\right) \leq \frac{64 \cdot \log^2(p)}{p}.$$

Assume that p is chosen so large that $64 \cdot \log^2(p)/p < 0.01$. Then letting

$$\Lambda_p = \left\{x \in I_0 : \sum_{h=1}^{2^p} X_{h,p}(x) > \frac{1}{2^2\log^2(p)}\right\},$$

we have $\mu(\Lambda_p) > 0.99$. We set $E_p = K_e^{(2^p)}$. For any $x \in \Lambda_p$ we have an $n \in [2^{2^p}, 2^{E_p}]$ such that

$$\frac{\mathbf{N}_n(f_p)(x)}{n} \geq \sum_{h=1}^{2^p} X_{h,p}(x) > \frac{1}{2^2\log^2(p)},$$

and

$$\frac{1}{p\log^2(p)} \leq \int_I f_p = 2^{-[p+\log(p)+\log(\log^2(p))]+1} \cdot 2^p \leq \frac{4}{p\log^2(p)}.$$

$\square$

Next we turn to the proof of Theorem 1.

Proof. By using Lemma 7 with $I = X = [0,1)$ choose p_0 and for each $p > p_0$ a p-block. Set $\phi_p = f_p / \int_I f_p$, and $\lambda_p = 1/(2^3 \log^2(p) \cdot \int_I f_p) \geq p/32$. By Lemma 7 for any $x \in \Lambda_p$ there is $n' \in [2^{2^p}, 2^{E_p}]$ such that

$$\frac{\mathbf{N}_{n'}(f_p)(x)}{n'} > \frac{1}{2^2 \log^2(p)}.$$

Now using the definition of $\mathbf{N}_{n'}$ we obtain

$$\frac{\mathbf{N}_{n'}(f_p)(x)}{n'} = \frac{\#\{k : f_p(T^k x)/k > 1/n'\}}{n'} = \frac{\#\{k : \phi_p(T^k x)/k > 1/(n' \int_I f_p)\}}{n'} <$$
$$\frac{\#\{k : \phi_p(T^k x)/k > 1/([n' \int_I f_p] + 1)\}}{[n' \int_I f_p] + 1} \cdot \frac{n' \int_I f_p + 1}{n'} =$$

using $n = [n' \int_I f_p] + 1$

$$\frac{\mathbf{N}_n(\phi_p)(x)}{n} \left(\int_I f_p + \frac{1}{n'} \right) < \frac{\mathbf{N}_n(\phi_p)(x)}{n} \cdot 2 \int_I f_p.$$

Hence for all x from a set of measure at least 0.99 there exist n such that $\mathbf{N}_n(\phi_p)(x)/n > \frac{1}{2^3 \log^2(p) \int_I f_p} = \lambda_p$.

Since $\lambda_p \cdot 0.99 \to \infty$ we have established Theorem 1. □

3 Proofs of Theorems 2 and 3

3.1 Proof of Theorem 2

Proof. Theorem 2 follows from Theorem 8 in [1]. It was shown there that for a sequence of nonnegative numbers c_n such that $\lim_n c_n/n = 0$ the following two statements are equivalent

1. $$\sup_n \frac{\#\{k : \frac{c_k}{k} > \frac{1}{n}\}}{n} < \infty;$$

   and

2. for all measure preserving systems $(Y, \mathcal{G}, \nu, S)$ and all $g \in L^1(\nu)$, the sequence $c_n \cdot g(S^n y)/n$ converges to zero ν a.e.

Taking the sequence $c_n = f(T^n x)$ for an ergodic transformation T shows that if the validity of the Return Time for the Tail Property in L^1 were to hold, then we should have for all $f \in L^1_+(\mu)$ for a.e. x,

$$\sup_n \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n} < \infty. \tag{18}$$

Condition (18) for all $f \in L^1_+(\mu)$ is equivalent to saying that

$$\sup_{\alpha > 0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} < \infty$$

for all $f \in L^1_+(\mu)$ for a.e. x. Consider an enumeration of the positive rational numbers r_k and define for each k the function $\mathbf{T}_k(f)(x) = \frac{\mathbf{N}_{r_k}(f)(x)}{r_k}$. We have

$$\sup_{\alpha>0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} = \sup_k \mathbf{T}_k(f)(x)$$

When T is ergodic it commutes with the family of powers of T. By the ergodic theorem this family is mixing. Indeed, we have

$$\lim_N \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^n(B)) = \mu(A)\mu(B)$$

so for each $\rho \geq 1$ there exists a n such that $\mu(A \cap T^n(B)) \leq \rho\mu(A)\mu(B)$. For each $\gamma \geq 1$ we have

$$\sup_k \mathbf{T}_k(\gamma f)(x) = \gamma \sup_k \mathbf{T}_k(f)(x).$$

Thus we can apply Theorem 4 of [1] to conclude that there exists a finite positive constant C such that for all $f \in L^1_+$,

$$\mu\{x : \sup_k \mathbf{T}_k(f)(x) > 1\} \leq C \int f d\mu.$$

This means that

$$\mu \left\{ x : \sup_{\alpha>0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} > 1 \right\} \leq C \int f d\mu.$$

Replacing the function f by f/λ provides a maximal inequality for the maximal function

$$\sup_{\alpha>0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha}.$$

From this we obtain easily a maximal inequality with the same constant C for

$$\sup_{n} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n}.$$

Having this constant for one ergodic transformation provides the same constant for all ergodic transformations. The ergodic decomposition would then show that

$$\sup_{(X,\mathcal{B},\mu,T)} \sup_{\|f\|_1=1} \sup_{\lambda>0} \lambda \cdot \mu\{x : \sup_{n} \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\} \le C < \infty.$$

This would contradict Theorem 1. $\square$

3.2 Proof of Theorem 3

Proof. Theorem 3 also follows from Theorem 1. We can argue also by contradiction. If we had the validity of the Return Times for Pairs property for (L^1, L^1) spaces then we would have the convergence in the universal sense of the averages

$$\frac{1}{N} \sum_{n=1}^{N} f(T^n x) \cdot g(S^n y) = \frac{\sigma_N}{N}$$

for $g \in L^1(\nu)$. This would imply the convergence to zero of

$$\frac{\sigma_N}{N} - \frac{\sigma_{N-1}}{N-1} = \frac{f(T^N x) g(S^N y)}{N} + \frac{\sigma_{N-1}}{N-1} \cdot \frac{N-1}{N} - \frac{\sigma_{N-1}}{N-1}.$$

This in turn would give the validity of the Return Time for the Tail property in L^1, but this was disproved in Theorem 2. $\square$

4 The counting problem and Birkhoff's theorem

Theorem 1 also helps to refine connections between Birkhoff's pointwise ergodic theorem and the counting problem. It provides an example of a maximal operator which is of restricted weak type (1,1) but does not satisfy a weak type (1, 1) inequality. However, this operator coincides with the one sided Hardy–Littlewood maximal function on characteristic functions of measurable sets. Let us see how and why.

One way to prove Birkhoff's pointwise ergodic theorem is via the maximal inequality

$$\mu\left\{x : \sup_N \frac{1}{N} \sum_{n=1}^{N} |f|(T^n x) > \lambda\right\} \leq \frac{1}{\lambda} \|f\|_1.$$

It turns out (see [7] for instance) that this maximal inequality is equivalent to the weak type (1,1) inequality for the Hardy–Littlewood maximal function on $\mathbb{T}$, the unit circle, that we identify with the interval $[-\frac{1}{2}, \frac{1}{2})$,

$$H(f)(x) = \sup_{t>0} \frac{1}{t} \int_0^t |f(x-y)| dy.$$

The following maximal function was introduced by the first author

$$A(f)(x) = \sup_{\lambda>0} \lambda \cdot m \left\{ 0 < y < x : \frac{|f(x-y)|}{y} > \lambda \right\}.$$

The interest in the operator A lies in the following results

1. It was used in [12] to give the details of the fact that the return time for the tail in all L^p spaces $1 < p \leq \infty$ is equivalent to the validity of Birkhoff's theorem in all L^r spaces for $1 < r \leq \infty$. In other words, the finiteness of $\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ shown in [1] is equivalent to Birkhoff's theorem in L^p for $1 < p \leq \infty$.

2. If one considers the characteristic function of a measurable set B, then simple computations show that

$$A(\mathbf{1}_B)(x) = H(\mathbf{1}_B)(x). \tag{19}$$

Thus the operator A satisfies a restricted weak type (1, 1) inequality in the sense that we have for all $\lambda > 0$

$$m\{x : A(\mathbf{1}_B)(x) > \lambda\} \leq \frac{1}{\lambda} m(B)$$

i.e. a weak type (1, 1) inequality for characteristic functions of measurable sets. (See also [15] or [7] for instance for more on restricted weak type inequalities.)

3. The operator A can be viewed as a continuous analog of the counting function studied in the previous sections. Furthermore, we have the following lemma.

Lemma 8. *Given* p, $1 \leq p \leq \infty$ *the following statements are equivalent*

(a) There exists a finite constant C *such that for all* $\lambda > 0$ *and* $(a_n) \in l^p(\mathbb{Z})$

$$\#\left\{i \in \mathbb{Z} : \sup_n \left(\frac{\#\{k > 0 : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n}\right) > \lambda\right\} \leq \frac{C}{\lambda^p}\|(a_n)\|_p^p. \tag{20}$$

(b) There exists a finite constant C *such that for all* $f \in L^p(\mathbb{T})$ *and* $\lambda > 0$ *we have*

$$m\{x : A(f)(x) > \lambda\} \leq \frac{C}{\lambda^p}\int |f|^p dm.$$

(c) We can find a finite constant C *such that for all* $f \in L^p_+(\mu)$ *for all measure preserving systems* $(X, \mathcal{B}, \mu, T)$

$$\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} \leq \frac{C}{\lambda^p}\int |f|^p d\mu$$

Proof. The proof uses known methods in ergodic theory such as transference or Rohlin's tower lemma. Details of such computations can be seen in [12]. So we only sketch some of them. We remark that (a) is equivalent to the following inequality.

There exists a finite constant C such that for all $\lambda > 0$, $(a_n) \in l^p(\mathbb{Z})$, positive integers K and I,

$$\#\left\{ i \in [-I, I] : \sup_{n \le K} \left(\frac{\#\{k > 0 : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n} \right) > \lambda \right\} \le \frac{C}{\lambda^p} \|(a_n)\|_p^p. \tag{21}$$

In order to prove that (a) and (b) are equivalent we use step functions of the form $f = \sum_{j=-I}^{I-1} a_j \mathbf{1}_{I_j}$ where $a_j \in \mathbb{R}$ and $a_j = 0$ for $|j| > I$. The interval I_i equals the dyadic interval $[\frac{i}{2^I}, \frac{i+1}{2^I})$.

To show that (a) and (c) are equivalent we use Rohlin's tower lemma where the tower is symmetric and of height $2J + 1$. Rohlin's lemma tells us that for any $\epsilon > 0$ and $J \in \mathbb{N}$ we can find disjoint sets $T^{-i}B$ for $-J \le i \le J$, such that the tower $\cup_{i=-J}^{J} T^{-i}(B)$ has total measure greater than $1 - \epsilon$. We take a function $f = \sum_{i=-J}^{J} a_i \mathbf{1}_{T^iB}$ and note that

$$\begin{aligned} \frac{\mathbf{N}_n(f)(x)}{n} &= \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n} \\ &\ge \sum_{i=-J}^{J} \mathbf{1}_{T^iB}(x) \frac{\#\{k \le J - |i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n}. \end{aligned}$$

Thus, the inequality

$$\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} \leq \frac{C}{\lambda^p}\int |f|^p d\mu$$

implies

$$\begin{aligned}\sum_{i=-J}^{J} \mu\left\{x \in T^i B : \sup_n \left(\frac{\#\{k \leq J-|i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n}\right) > \lambda\right\} \\ \leq \frac{C}{\lambda^p}\mu(B)\sum_{i=-J}^{J}|a_i|^p.\end{aligned} \tag{22}$$

As (22) equals

$$\mu(B)\cdot \#\left\{-J \leq i \leq J : \sup_n \left(\frac{\#\{k \leq J-|i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n}\right) > \lambda\right\}$$

we have

$$\begin{aligned}&\#\left\{i \in \mathbb{Z} : \sup_{0<n\leq K} \left(\frac{\#\{k > 0 : \frac{a_{k+i}}{k} \geq \frac{1}{n}\}}{n}\right) > \lambda\right\} \\ &\leq \lim_J \#\left\{-J \leq i \leq J : \sup_{0<n\leq K} \left(\frac{\#\{k \leq J-|i| : \frac{a_{k+i}}{k} \geq \frac{1}{n}\}}{n}\right) > \lambda\right\}.\end{aligned}$$

$\square$

So Theorem 1 gives us the following contribution to the problem of characterizing operators for which a restricted weak type (1,1) inequality implies a weak type (1,1) inequality. (See [7] for more on this problem.) The operator A does not satisfy a weak type (1,1) inequality. It is shown in [7] that if an operator is generated by convolutions, then a restricted weak type (1,1) inequality implies a weak type (1,1) inequality. Such is the case of the Hilbert transform and the Hardy–Littlewood maximal function.

Next we list some of the properties of the operator A.

Theorem 9. *The operator A defined on $\mathbb{T}$ by the formula*

$$A(f)(x) = \sup_{\lambda>0} \lambda \cdot m \left\{ 0 < y < x : \frac{|f(x-y)|}{y} > \lambda \right\}$$

has the following properties

1. *It coincides with the one sided Hardy–Littlewood maximal function when f is the characteristic function of a measurable set on $\mathbb{T}$ hence it satisfies a restricted weak type (1,1) inequality.*

2. *It maps functions in L^p to functions in weak L^p.*

3. *There exists a positive function $f \in L^1(\mathbb{T})$ such that $A(f)(x) \not< \infty$ for a.e. x in $\mathbb{T}$.*

Proof. Statements (1) and (2) follow from Lemma 8.

The last statement is a consequence of Theorem 1. The arguments developped in [1] (cf. Theorem 4) indicate that if we had $A(f)(x) < \infty$ for a.e. x then we would have a weak type (1,1) inequality for A. By Lemma 8 this would imply a weak type (1,1) inequality for $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, a conclusion that we disproved in Theorem 1. $\square$

References

[1] I. Assani, "Strong Laws for weighted sums of iid random variables," *Duke Math J.*, 88, 2, (1997), 217-246.

[2] I. Assani, "Convergence of the p-Series for stationary sequences, *New York J. Math.,* 3A, (1997), 15-30.

[3] I. Assani, "A weighted pointwise ergodic theorem," *Ann. Inst. Henri Poincare,* 34, 1, (1998), 139-150.

[4] J. Bourgain, "Return Time sequences of dynamical systems," IHES, Preprint, (1988).

[5] J. Bourgain, "Temps de retour pour des systems dynamiques," *C.R. Acad. Sci. Paris*, t. 306, Série I, (1988), 483-485.

[6] J. Bourgain, "Pointwise ergodic theorems for arithmetic sets", With an appendix by the author, Harry Fürstenberg, Yitzhak Katznelson and Donald S. Ornstein, *Inst. Hautes Études Sci. Publ. Math.* No. 69 (1989), 5–45.

[7] C. Bennett and R. Sharpley, *Interpolation of operators,* Academic Press, New York, 1988.

[8] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Fifth edition, Oxford University Press, (1979).

[9] R. Jones, J. Rosenblatt and M. Wierdl, *Counting in Ergodic Theory*, Cand. J. Math., 51, (1999), 996-1019.

[10] L. Kuipers and H. Niederreiter, *Uniform distribution of Sequences*, John Wiley and Sons, (1974).

[11] J. Lamperti, *Probability: A survey of the Mathematical Theory*, W. A. Benjamin Inc., New York (1966).

[12] K. Noonan, *Return Times for the tail and Birkhoff's theorem,* Master's thesis, UNC Chapel Hill. Dec. 2002.

[13] D. Rudolph, "A joining proof of Bourgain's return time theorem, *Erg. Th. and Dyn. Syst.*, 14, (1994), 197-203.

[14] D. Rudolph, "Fully generic sequences and a multiple term return times theorem" *Invent. Math.* , 131, (1998), 199-228.

[15] E. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces,* Princeton University Press, 1971.